Spring 2020

Designing a Reporter Construct and CRISPR Construct to Identify the Role of AtCDT1 in Aluminum and Cadmium Tolerance in Arabidopsis thaliana

Hannah McCurry
hannah.mccurry@trincoll.edu

TRINITY COLLEGE

DESIGNING A REPORTER CONSTRUCT AND CRISPR CONSTRUCT TO IDENTIFY THE ROLE OF *AtCDT1* IN ALUMINUM AND CADMIUM TOLERANCE IN *ARABIDOPSIS THALIANA*

BY

HANNAH R. MCCURRY

A THESIS SUBMITTED TO
THE FACULTY OF THE DEPARTMENT OF BIOLOGY
IN CANDIDACY FOR THE BACCALAUREATE DEGREE
WITH HONORS IN BIOLOGY

DEPARTMENT OF BIOLOGY

HARTFORD, CONNECTICUT

MAY 6 2020

DESIGNING A REPORTER CONSTRUCT AND CRISPR CONSTRUCT TO IDENTIFY THE ROLE OF *AtCDT1* IN ALUMINUM AND CADMIUM TOLERANCE IN *ARABIDOPSIS THALIANA*

BY

HANNAH R. MCCURRY

Honors Thesis Committee

Approved:

_____
Susan M. Bush, Advisor

_____
Lisa-Anne A. Foster

_____
Robert J. Fleming

_____
Claire T. Fournier

Date: _____

TABLE OF CONTENTS

Acknowledgements

I would like to thank Professor Susan Bush for letting me work in her lab and for allowing me the opportunity to write this thesis. She has provided immense help and advice both in and outside of the lab. Without Professor Bush, I would not have had this amazing experience, nor would I know that plant biology is what I would like to pursue in my future career. I thank Professor Robert Fleming, Professor Lisa-Anne Foster, and Professor Claire Fournier for providing feedback along the way in writing this thesis. Thank you to the whole Trinity College Biology Department for allowing this research and thesis to be possible. I would also like to thank Dr. Mily Ron from the University of California, Davis for providing base plasmids and a detailed protocol used in this research.

I'd like to thank my family for supporting me throughout college and certainly throughout my senior year in writing this thesis. I would also like to thank my best friend Mary Tursi '20, for reassuring me every step of the way that I would indeed finish this thesis and for being an amazing friend throughout these past four years, a fact that certainly will not change long into the future.

Abstract

The gene *CDT1* was first discovered in the crabgrass *Digitaria cilaris* to confer cadmium tolerance. Five homologs of this gene exist in rice (*Oryza sativa*), one of which, *OsCDT3*, was found to confer aluminum tolerance, instead of cadmium tolerance. The homolog of this gene in *Arabidopsis thaliana* (*AtCDT1*), is currently under study. Arabidopsis plants carrying a T-DNA insertion mutation in this gene were found to overexpress *AtCDT1* in response to cadmium, but not to aluminum, suggesting that *AtCDT1* is cadmium responsive, not aluminum responsive. This research study aims to learn more about the *AtCDT1* gene and its potential involvement in tolerance to aluminum toxicity. The current study proposes a reporter construct to show where in the plant and when during development the *AtCDT1* gene is expressed in Arabidopsis, with predictions that it is expressed in the cell membrane of cells in the roots, and throughout development of the Arabidopsis plant. Preliminary CRISPR constructs were designed to be used in further research to design a knockout line of Arabidopsis for the *AtCDT1* gene, to determine its potential function in tolerance to only aluminum, only cadmium, both, or neither of these metals.

Introduction

Aluminum Tolerance

Soils naturally contain certain metals, such as aluminum, that can be toxic to many plants, particularly in low pH environments. Low pH levels activate metals such as aluminum in the soil, which inhibit the root growth of plants living in these environments (Gazey 2018). Without proper root growth, the plant lacks in its ability to absorb nutrients to sustain growth, resulting in poor crop growth and yield (Gazey 2018). Methods used by plants to fight aluminum toxicity include absorption and storage of aluminum within the plant cell vacuoles, to prevent its degrading effect on the roots, primarily using ATP-binding cassette (ABC) transporters that use ATP to transport aluminum across the vacuolar membrane (Eekhout et al. 2017). Another method used to tolerate aluminum toxicity is chelation, whereby the plant secretes organic molecules such as malate or citrate that bind to aluminum ions and form harmless compounds, preventing their inhibitory nature (Heck & Bailey 1950).

A number of genes have been identified that are related to aluminum tolerance, such as *ALMT1,* which codes for a protein that transports malate and is important in the chelation method of aluminum tolerance (Hoekenga et al. 2006); *ALS1*, which codes for an ABC transporter-like protein (Larsen et al. 2005); and *STOP1*, which regulates *ALMT1* and *ALS1* (Iuchi et al. 2007). By studying aluminum tolerance at the molecular level, it is possible to discover more genes that are involved in aluminum tolerance. These genes could potentially confer aluminum tolerance to plants that are sensitive to aluminum toxicity, allowing them to grow in low pH soils with high aluminum content, thereby allowing for increased crop yield of these aluminum sensitive plants.

Cadmium Tolerance and the CDT1 *gene*

The *cadmium tolerance gene 1* (*CDT1*) gene has recently been found to be involved in aluminum tolerance in rice (*Oryza sativa*), even though this gene was initially found through its involvement in cadmium tolerance in the crabgrass species *Digitaria cilaris* (Xia et al. 2013; Kuramata et al. 2008). Cadmium ions in the soil are toxic to plants at low pH, similarly to aluminum ions, and study of this toxicity has revealed a number of mechanisms and genes involved in conferring tolerance to cadmium. These mechanisms of cadmium tolerance are similar to those of aluminum tolerance, including sequestration of cadmium in the vacuoles, and chelation of cadmium ions in the soil, after which the harmless cadmium-complexes are stored in the vacuole (reviewed in Clemens 2001). Another mechanism involves transporting cadmium to be stored in parts of the plant that are less susceptible to cadmium toxicity, *i.e*. the shoots of the plant rather than the roots (Clemens 2001).

Kuramata et al. (2008) first isolated a 521 base pair long cDNA clone of *DcCDT1* by screening for *D. cilaris* cDNA that induced cadmium tolerance when transformed into mutant yeast cells that were hypersensitive to cadmium. The resulting peptide, DcCDT1, is cysteine rich, with 15 cysteines out of 55 total amino acids. Expression studies performed by Kuramata et al. (2008) suggested that DcCDT1 is localized in the cell membrane. Transgenic expression of *DcCDT1* in *Arabidopsis thaliana* increased tolerance to cadmium. Homologs of *DcCDT1* were found in multiple other crops, notably in rice (*OsCDT1*) and Arabidopsis (*AtCDT1*) (Kuramata et al. 2008).

Rice has five homologs of the *CDT1* gene (*OsCDT1-5)*, and their individual characteristics and functions were studied by Xia et al. (2013). The first homolog, *OsCDT1*, improves cadmium tolerance when expressed in Arabidopsis and yeast lines, which is

consistent with the research previously done by Kuramata et al. (2008) with *DcCDT1* (Xia et al. 2013). The third rice *CDT1* homolog, *OsCDT3,* was surprisingly found to aid in aluminum tolerance, but not in cadmium tolerance. Through expression studies conducted by Xia et al. (2013), OsCDT3 was found to be a membrane-bound protein that is induced by aluminum and prevents aluminum entrance into the cell by binding to it. These expression studies uphold the previous research surrounding the localization of OsCDT1 done by Kuramata et al. (2008). This information was used to predict the cellular localization of the AtCDT1 protein, as also being a plasma membrane protein involved in binding to aluminum or cadmium (Figure 1).

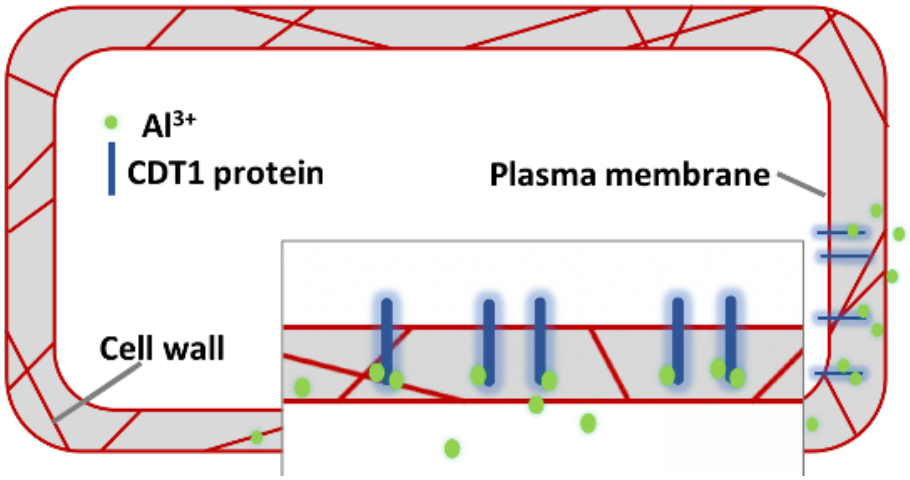

Figure 1: AtCDT1 binds aluminum ions in the cell wall, preventing their entrance into the cell, and inhibiting their toxic effects.

Kuramata et al. (2008) identified two open reading frames (ORFs) for *AtCDT1* that overlap but produce two different peptides. These two open reading frames were labeled ORF3 and ORF2 and were reported on The Arabidopsis Information Resource (TAIR) website. The third open reading frame (ORF3) has no amino acid sequence similarity to DcCDT1, while the second (ORF2) holds high correlation to DcCDT1, and has a cysteine-rich region, which is present in both the rice and *D. cilaris* homologs of this gene (Xia et al. 2013; Kuramata et al. 2008). This finding was recently upheld when the amino acid sequence

of the expressed transcript of *AtCDT1* was published on TAIR: the expressed transcript matches ORF2, not ORF3, showing that ORF2 produces the AtCDT1 peptide that is homologous to DcCDT1 and OsCDT3 (Figure 2).

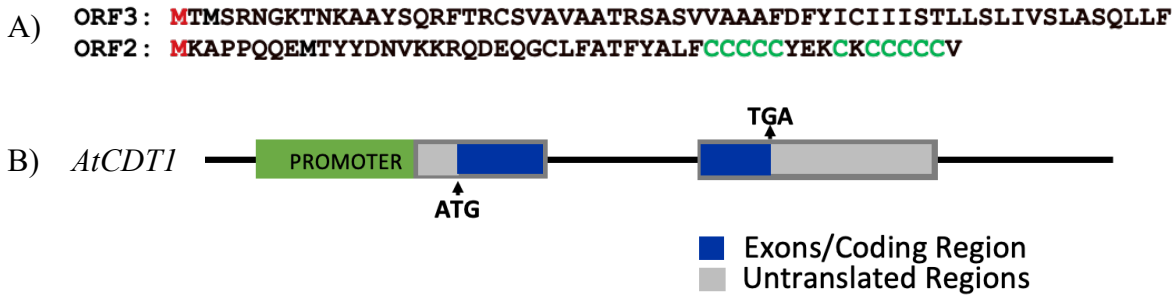

Figure 2: A) Amino acid sequences of the two proposed open reading frames (ORFs) of the AtCDT1 protein. The second open reading frame (ORF2), holds sequence similarity to the original DcCDT1 protein, and is cysteine rich (shown in green), as seen in the rice and *D. cilaris* homologs. B) The Arabidopsis *CDT1* gene, showing the expressed reading frame with two exons (blue), one intron (black line), and upstream and downstream untranslated regions (gray). The start codon (ATG) and stop codon (TGA) are shown.

In Arabidopsis, the cysteine-rich transmembrane module (CYSTM) is a loosely related protein family, whose members are important in plant response to environmental stressors. The Arabidopsis analog of CDT1 was found to be a part of this protein family (Xu et al. 2017), providing evidence that the AtCDT1 protein is indeed important in Arabidopsis response to environmental stress, such as aluminum and cadmium toxicity in the soil.

The exact function of the Arabidopsis *CDT1* gene (*AtCDT1*) is still being studied, especially in terms of its response to aluminum, given the interesting involvement of *OsCDT3* in aluminum tolerance in rice. *AtCDT1* has homology to both *DcCDT1* and *OsCDT3,* giving the possibility of *AtCDT1* to be important in either cadmium or aluminum tolerance. Analysis of two Arabidopsis *CDT1* insertional mutants in the promoter of *AtCDT1* (mutants *cdt1-2* and *cdt1-3*) showed no significant difference in root length in response to aluminum or cadmium between the mutant plants and Arabidopsis plants of wild type line

Columbia (Col) (Michalopoulos et al. 2018). These mutants were found to lead to an overexpression of *AtCDT1* in response to cadmium treatment, but no such overexpression was seen in response to aluminum (Michalopoulos; Figure 3).

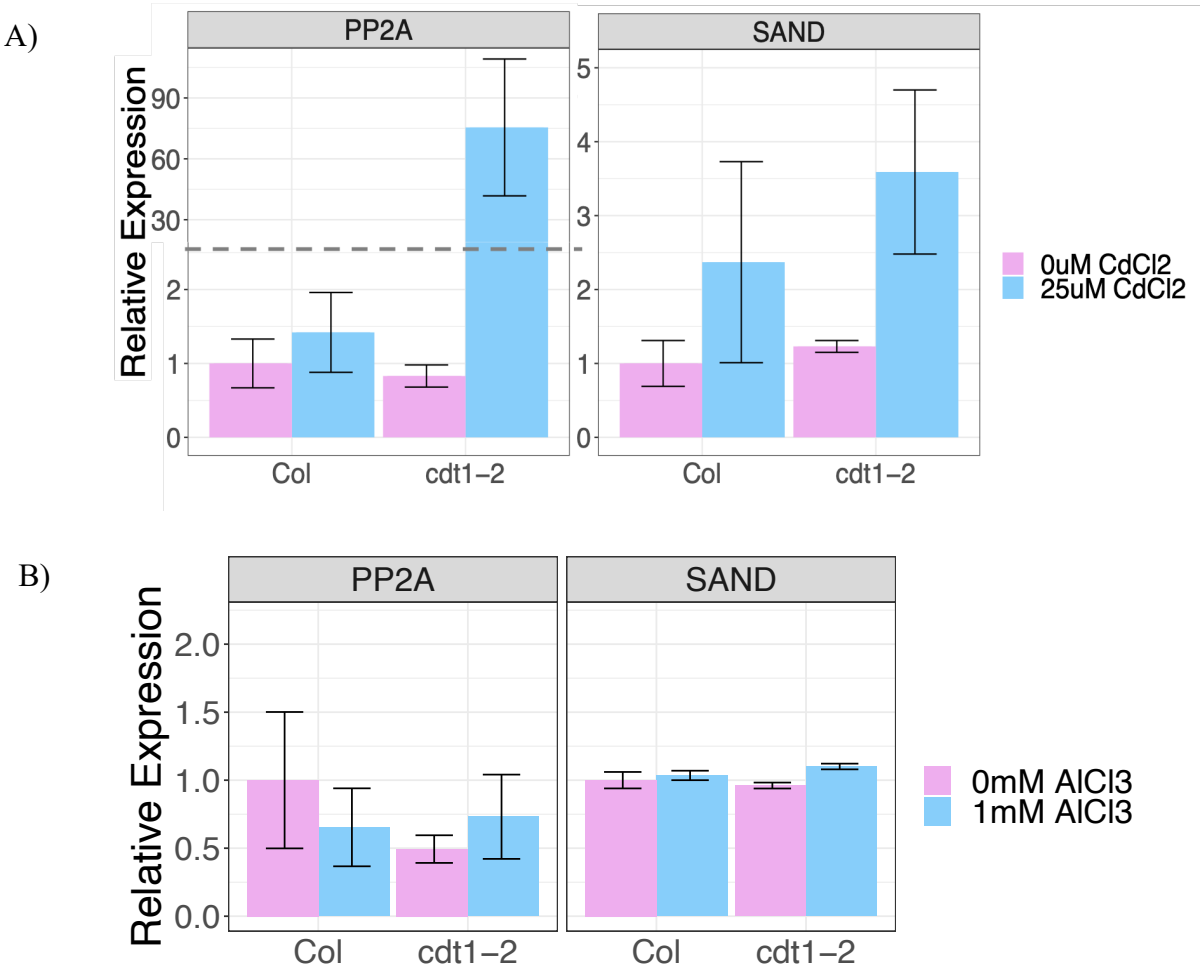

Figure 3: Relative expression of *AtCDT1* in wild type (Col) and *cdt1-2* mutant Arabidopsis plants in response to A) cadmium and B) aluminum. The reference genes used, *PP2A* and *SAND* , are expressed at a consistent rate in Arabidopsis, and *AtCDT1* expression was determined based off of the expression of these genes (from Michalopoulos et al. 2018).

Whether *AtCDT1* is involved in only cadmium tolerance (as with the *D. cilaris* homolog *DcDCT1*), only aluminum tolerance (as with the rice homolog *OsCDT3*), or in tolerance to both or neither of these metals is currently under study. This research aims to help answer this question.

The CRISPR/Cas9 Mechanism

To learn more about the specific function of this gene, the CRISPR/Cas9 mechanism will be used to create a knockout of *AtCDT1*. CRISPR is a natural mechanism found in bacteria, where it is used to fight viral infection (Broad Institute, 2018). When a bacterium is infected by a virus and the viral DNA begins to propagate, the bacterium will create a spacer sequence that is complementary to this viral DNA. This spacer sequence is integrated into the bacterium's own DNA. The spacer sequence is transcribed, and the resulting RNA is used to guide associated proteins to the viral DNA, where it makes double stranded cuts in the DNA sequence, thereby functionally destroying the viral DNA, and preventing further viral infection. The sequence of bacterial DNA that is made up of multiple spacer sequences from multiple viral infections, interspaced with repeating sequences to separate each spacer sequence from another, is called the CRISPR sequence (hence its name: Clustered Regularly Interspaced Short Palindromic Repeats).

This natural process used in association with the Cas9 endonuclease protein (CRISPR associated protein 9), allows for CRISPR to contribute heavily to the field of molecular biology (Broad Institute 2019). A researcher may design a transgenic vector including the *Cas9* gene sequence and a sequence that is complementary to the gene of interest in the study (which acts as the spacer sequence), termed the protospacer sequence. The transgenic sequences are integrated into the host organism as T-DNA insertions. The protospacer sequence is transcribed into a guide RNA (gRNA), which associates with the transcribed and translated Cas9 protein, and guides the entire mechanism to the gene of interest, where it makes a double stranded cut in the DNA (Figure 4). Appended to the protospacer sequence is the protospacer adjacent motif (PAM) sequence, which is necessary for proper function of

the Cas9 protein cleaving the DNA at the target sequence. The cell will try to fix its broken DNA in one of two ways: through non-homologous end joining (NHEJ) or through homology driven repair (HDR). NHEJ often introduces mutations such as insertions or deletions into the gene of interest, and thereby can functionally destroy the gene, whereas HDR allows for the possibility of an externally provided sequence to be introduced into the gene of interest. In this research, the CRISPR/Cas9 mechanism will be used with two gRNAs, specific to a point upstream of the *AtCDT1* gene and one at the end of with the first exon of *AtCDT1,* with the specific result of NHEJ in the *AtCDT1* gene in order to design a *CDT1*-knockout line of Arabidopsis.

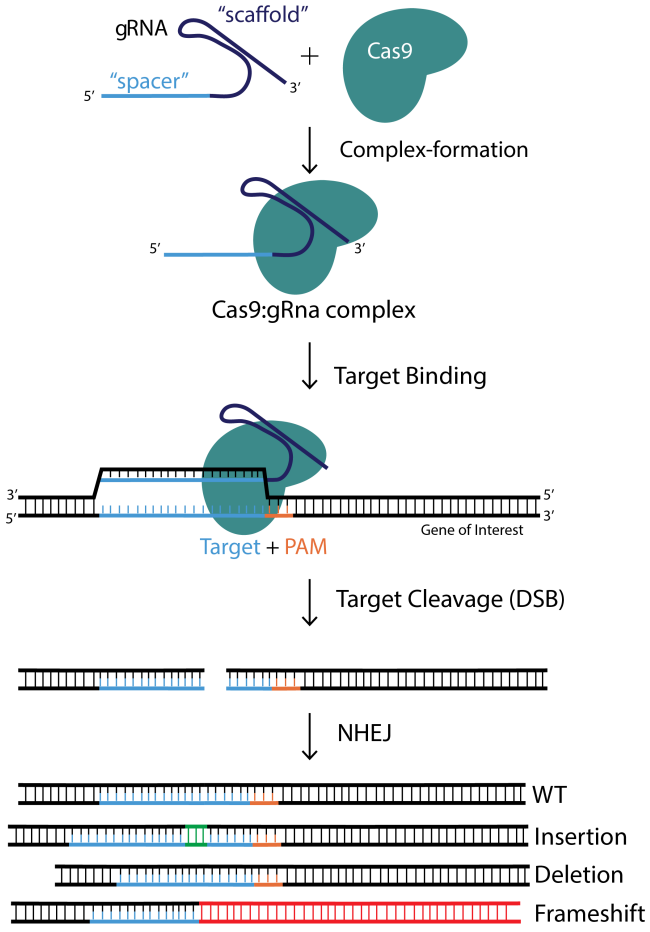

Figure 4: Visual representation of the CRISPR/Cas9 mechanism with the specific result of non-homologous end joining (NHEJ). The "spacer" sequence is the sequence that is complementary to the target gene sequence, which will be cut by Cas9, creating a double stranded break (DSB). The cell will then try to fix its broken DNA using NHEJ, which introduces mutations in the gene sequence. (Figure modified from Vector Biolabs)

Goals

This research project aims to create a reporter line and a knockout model for the *AtCDT1* gene, in order to learn more about this gene in Arabidopsis. The visual marker *GUS* will be used to determine where in the plant *AtCDT1* is expressed and when during development the gene is expressed. This *GUS* assay will provide insight into the expression patter of *AtCDT1*, which will be compared to the predicted expression pattern based on the *OsCDT3* homolog from rice. The CRISPR/Cas9 mechanism will be used to generate a knockout of the *AtCDT1* gene. By deleting a portion of this gene in the Arabidopsis genome, its function can be determined by comparison of the root length of these knockout plants to the root length of wildtype plants, in terms of its tolerance to aluminum and/or cadmium.

Materials and Methods

Growth Conditions

Arabidopsis thaliana seedlings of the wild type accession Columbia (Col-0) were planted and kept in a growth chamber set at 23° Celsius under constant light and were watered approximately every three days (Figure 5). Since the growth rate of roots of individual Arabidopsis plants would be measured, the variation seen in these plants was not expected to pose an issue in extrapolating results based on the effects of cadmium or aluminum stress.

Figure 5: Col-0 plants after two weeks of growth under constant light at 23° Celsius.

Creating the reporter construct pSB109

Escherichia coli (*E. coli*) stocks of the plasmid *pMDC164* were streaked onto Miller's LB agar plates with kanamycin. This plasmid contains the *GUS* reporter gene and the selectable marker of kanamycin resistance (Curtis & Grossniklaus 2003; Appendix 1). It was used as the base plasmid to build a vector containing the *AtCDT1* promoter directly next to a *GUS* reporter sequence (Figure 6). Overnight liquid cultures were prepared from the streak.

Minipreps of *pMDC164* were conducted to extract the plasmid DNA from the *E. coli* host using the Invitrogen Miniprep Kit. The plasmid DNA from these minipreps was used in subsequent reactions to create the GUS reporter construct.

PCR of *Col-0* Arabidopsis DNA was performed, with primers specific to the *AtCDT1* promoter sequence, that also contained recognition sites for *BamHI* (in the forward primer) and *SalI* (in the reverse primer). This PCR product of the *AtCDT1* promoter flanked by cut sites for *BamHI* and *SalI*, and the *pMDC164* vector, were digested overnight with the restriction enzymes *BamHI* and *SalI. BamHI* and *SalI* flank the negative selectable marker *ccdb* in *pMDC164.* This *ccdb* sequence, when transformed into *E. coli*, codes for a peptide

that kills *E. coli* cells. Digestion of *pMDC164* with *BamHI* and *SalI*, removes the *ccdb* sequence. The resulting piece of the *pMDC164* plasmid containing the *GUS* sequence, and the digested *AtCDT1* promoter were ligated together (Figure 6). The ligation reaction was performed with the following ratio of the *pMDC164* vector to the *AtCDT1* promoter fragments: three volumes of the *pMDC164* fragment were added for every one volume of the *AtCDT1* promoter fragment.

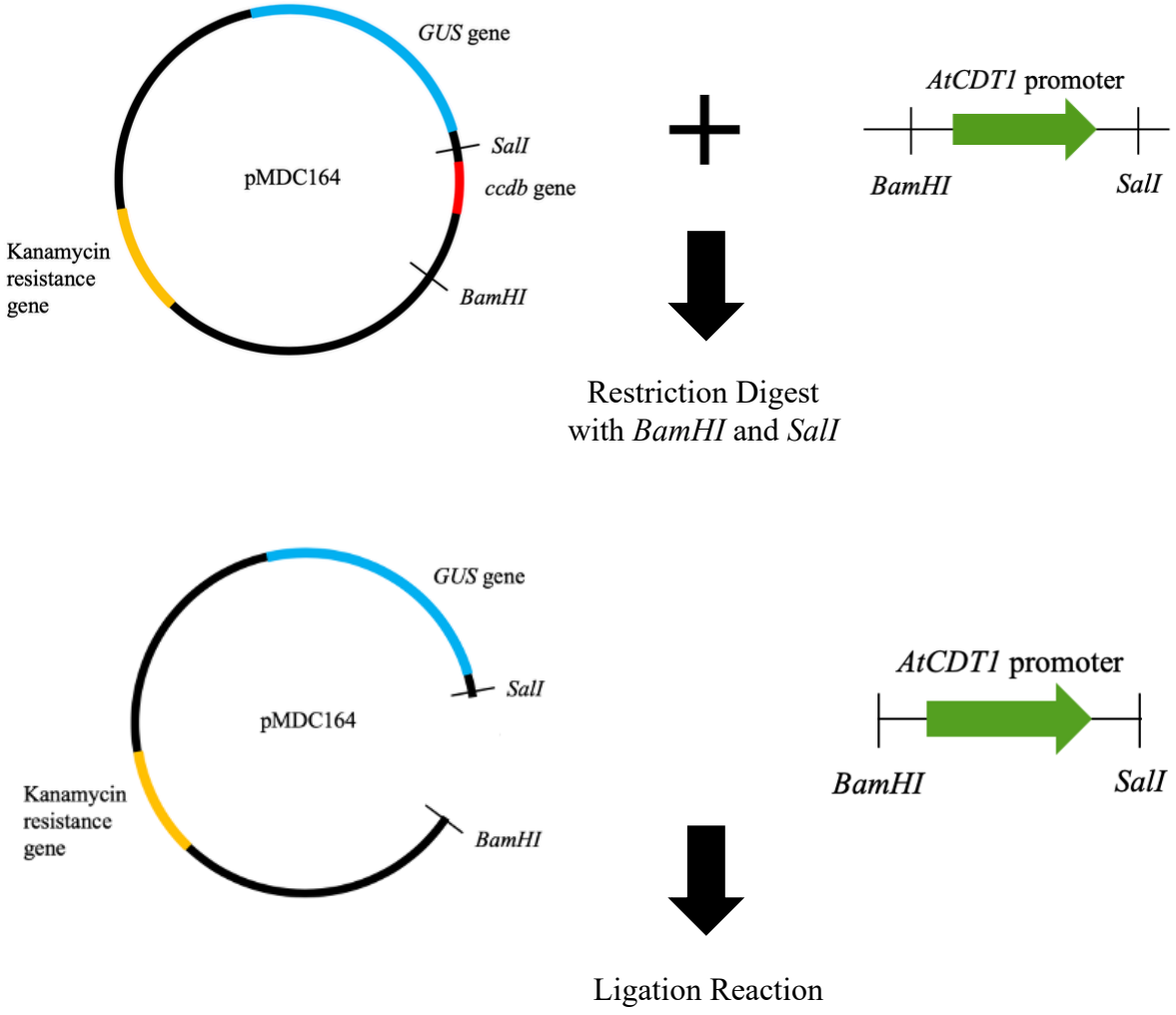

Figure 6: Diagram of the methods used to create the reporter construct, with a restriction digestion of both the base plasmid pMDC164 and the *AtCDT1* promoter sequence insert with *BamHI* and *SalI*, followed by a ligation of these two sequences together.

The resultant reporter construct, *pSB109* (Appendix 1), was transformed into competent DH5α *E. coli* cells, which were then plated on LB agar plates with kanamycin and left to grow overnight. Minipreps from overnight liquid cultures were performed to extract the vector DNA, which was sent for sequencing using the M13 forward primer by GeneWiz. The sequence was run through a BLAST analysis to confirm the presence of the *AtCDT1* promoter sequence in this plasmid (Figure 10).

Agrobacterium-mediated transformation

Plasmid DNA of *pSB109* was transformed into competent LBA4404 *Agrobacterium tumefaciens* cells using the gene electroporation system with a Gene-Pulser set to 25 μF, 200 Ω, and 2.5 kV and pulsed for five seconds. The transformed *A. tumefaciens* cells were plated on LB agar with kanamycin and streptomycin for 48 hours, and liquid cultures were prepared. Fifty mL of the liquid cultures were centrifuged at 3400 rpm for 10 minutes to pellet the cells, the supernatant was discarded, and 10 μL of inoculation medium (Appendix 2) was immediately added to resuspend the cells. Arabidopsis plants were infected with *A. tumefaciens* containing the reporter construct by pipetting this mixture in droplets, using filter pipette tips, directly onto the flowering buds of Arabidopsis plants and to the shoot apical meristems, following the modified floral dip method from Clough & Bent (1998). This process was repeated once more at five days after the first application. This direct application of the mixture containing the reporter construct to the flowers and meristematic tissue of the plants ensured that the reporter construct was integrated into the Arabidopsis genome of developing seeds and will be expressed in all tissues of the future plant.

Creating the CRISPR construct

Protospacer DNA oligomers that designate the Cas9 target sites were ordered, specifically to include overhangs that match the cut sites of the *BbsI* restriction enzyme, in both forward and reverse orientations. These protospacers were complementary to two sequences in the *AtCDT1* gene: upstream of the gene, and in the middle of exon 1 of the gene (Figure 7), and code for the two guide RNAs that would be used in the CRISPR/Cas9 mechanism. The forward and reverse protospacers of each type were annealed together creating double stranded inserts to be integrated ultimately into a vector containing the *Cas9* gene sequence.

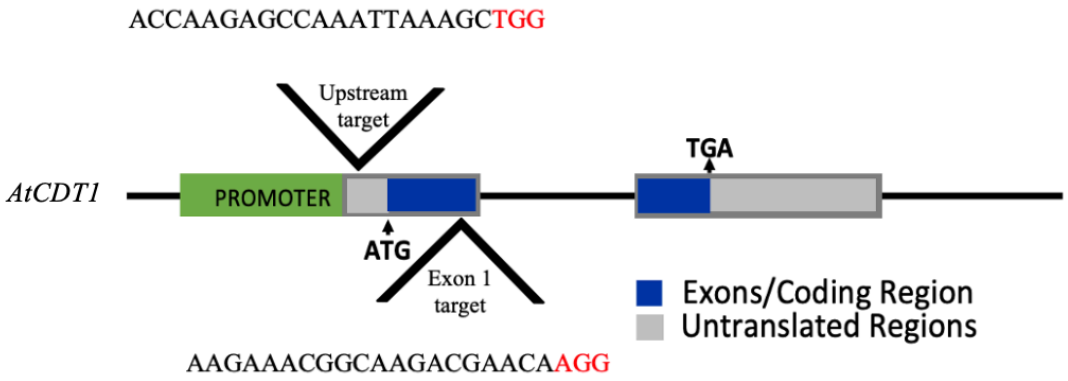

Figure 7: The *AtCDT1* gene structure with the upstream and exon 1 protospacer target sites, indicating where the Cas9 protein will make double stranded cuts in the DNA. The start codon (ATG) and the stop codon (TGA) are shown; the PAM sequences are denoted in red.

In order to integrate two guide RNAs into the CRISPR/Cas9 mechanism, the protocol outlined by Pauwels et al. (2018) and detailed by Ron (2018) was followed. The two protospacer sequences were ligated separately into two provided base plasmids, *pMR218* and *pMR299* (Ron 2018; Appendix 1). The protospacers were integrated into these base plasmids in the following orientation: the exon 1 protospacer was ligated into *pMR218*; and the upstream protospacer was ligated into *pMR299* (Figure 8).

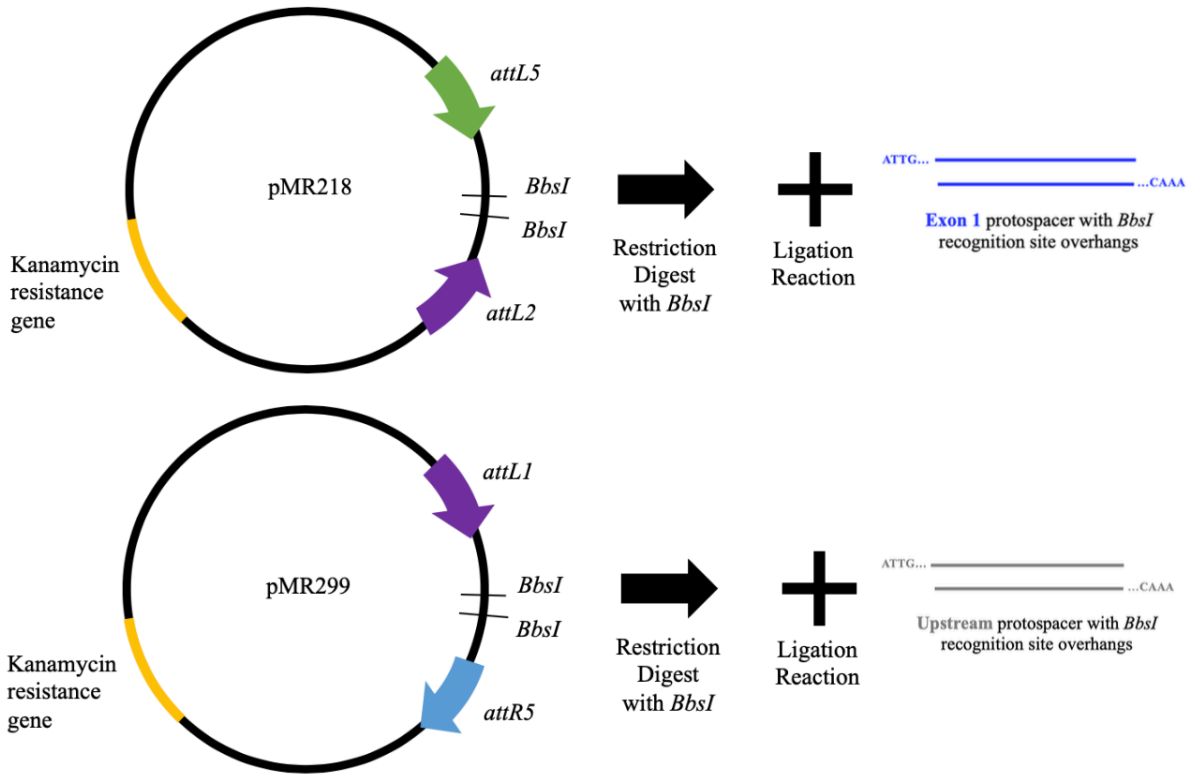

Figure 8: Diagram of the plasmids pMR218 and pMR299 with *BbsI* cut sites, and the protospacer sequences that target exon 1 of the *AtCDT1* gene and upstream of the *AtCDT1* gene, also with *BbsI* cut sites. The exon 1 protospacer was inserted into pMR218, while the upstream protospacer was inserted into pMR299.

Results

Reporter Construct

The reporter construct *pSB109* was created using restriction enzymes to digest the pMDC164 vector and the *AtCDT1* promoter insert, followed by a ligation reaction to attach these two segments together (Figure 9). The *pSB109* plasmid did not contain the negative selectable marker *ccdb,* and was therefore able to survive, unlike any *E. coli* cells containing an un-digested *pMDC164,* containing the *ccdb* sequence. The presence of the *AtCDT1* promoter sequence in the vector was confirmed using a BLAST alignment of the Sanger sequencing results of the plasmid. A portion of the Arabidopsis *AtCDT1* gene sequence was

found to align with 100% sequence similarity to the plasmid sequence, confirming the presence of the *AtCDT1* promoter next to the *GUS* sequence in *pSB109* (Figure 10). Furthermore, ten of the BLAST results align with Chromosome 1 of the Arabidopsis genome, the location of *AtCDT1*.

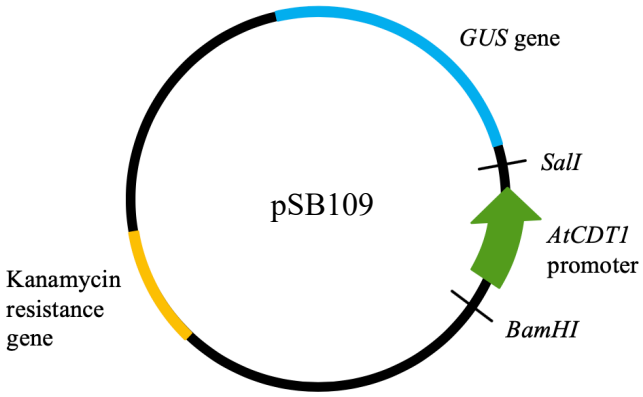

Figure 9: The pSB109 reporter construct, with the *GUS* reporter gene driven by the *AtCDT1* promoter. A kanamycin resistance selectable marker gene is included in this construct.

Arabidopsis thaliana genome assembly, chromosome: 1	1153	1153	89%	0.0	100.00%	LR699770.1
Arabidopsis thaliana genome assembly, chromosome: 1	1153	1153	89%	0.0	100.00%	LR699765.1
Arabidopsis thaliana genome assembly, chromosome: 1	1153	1153	89%	0.0	100.00%	LR699755.1
Arabidopsis thaliana genome assembly, chromosome: 1	1153	1153	89%	0.0	100.00%	LR215052.1
Arabidopsis thaliana chromosome 1 sequence	1153	1153	89%	0.0	100.00%	CP002684.1
Arabidopsis thaliana chromosome 1 BAC F14G24 genomic sequence, complete sequence	1153	1153	89%	0.0	100.00%	AC019018.9
Arabidopsis thaliana At1g52830 gene, 3' UTR	1153	1153	89%	0.0	100.00%	AY785464.1
Arabidopsis thaliana genome assembly, chromosome: 1	1147	1147	89%	0.0	99.84%	LR699745.2
Arabidopsis thaliana genome assembly, chromosome: 1	1147	1147	89%	0.0	99.84%	LR699750.1
Arabidopsis thaliana genome assembly, chromosome: 1	1134	1134	89%	0.0	99.52%	LR699760.1
PREDICTED: Arabidopsis lyrata subsp. lyrata auxin-responsive protein IAA6 (LOC9330465), mRNA	553	553	60%	5e-153	90.34%	XM_002894358.2
Arabidopsis thaliana cadmium tolerance 1 (CDT1), mRNA	298	298	23%	2e-76	100.00%	NM_202281.4
Arabidopsis thaliana clone 98934 mRNA sequence	110	110	8%	1e-19	98.39%	DQ108901.1
Expression vector pFJ1.1-yk1350a08, complete sequence	97.1	97.1	9%	1e-15	93.94%	FJ602701.1
Expression vector unc-68:GFP(1-8), complete sequence	95.3	95.3	7%	4e-15	100.00%	AF097552.1
Expression vector unc-68:GFP, complete sequence	95.3	95.3	7%	4e-15	100.00%	AF097551.1
Salmonella enterica subsp. enterica serovar Infantis strain TR01 Mutant chromosome	93.5	93.5	7%	1e-14	98.15%	CP040601.1
Stx converting phage vB_EcoS_P22, complete genome	93.5	93.5	7%	1e-14	98.15%	KU238069.1
Stx converting phage vB_EcoS_P32, complete genome	93.5	93.5	7%	1e-14	98.15%	KU238068.1
Stx converting phage vB_EcoS_P27, complete genome	93.5	93.5	7%	1e-14	98.15%	KU238067.1

Figure 10: BLAST results showing reporter construct sequence homology to *AtCDT1*, confirming the presence of the *AtCDT1* promoter in *pSB109*.

Upon integration into the Arabidopsis genome, the *GUS* gene would be expressed, and blue color would be seen, wherever in the plant and whenever during development

AtCDT1 is expressed. Arabidopsis seeds containing the *GUS* gene driven by the *AtCDT1* promoter have been harvested, but the *GUS* assay to determine where and when in the plant *AtCDT1* is expressed will not occur this semester.

CRISPR Construct

The two protospacer sequences, coding for guide RNAs to target regions upstream of the *AtCDT1* gene and in exon one of the gene, were used in a ligation reaction to integrate each sequence individually into a base plasmid. The exon1 protospacer was integrated into the base plasmid *pMR218* and the upstream protospacer was integrated into the base plasmid *pMR299* (Figure 11). In the future, these preliminary constructs will be used in a reaction involving homologous recombination to orient the protospacer sequences next to each other into the same plasmid, and then integrate both into a Cas9 plasmid, to create the final CRISPR construct.

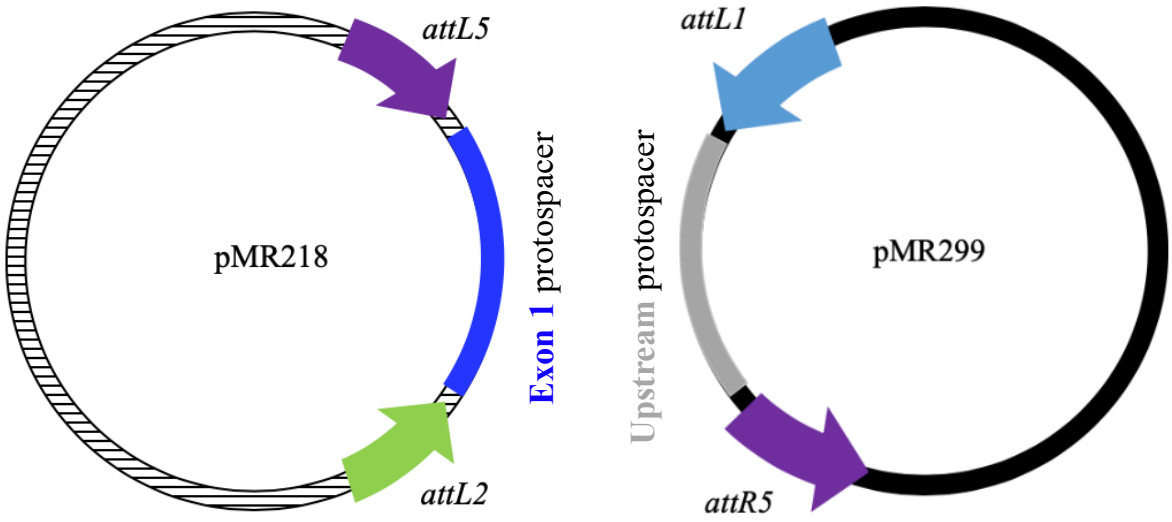

Figure 11: The two preliminary constructs containing the protospacer sequences separately in different plasmids, to be used in a homologous recombination reaction to orient the protospacer sequences next to each other in the same plasmid.

Discussion

Future Steps

Arabidopsis seedlings carrying the sequences included in the reporter construct *pSB109* have been harvested and will be grown in the future. A *GUS* assay will be performed using the whole plant over the span of the plant's development to see where in the plant and when during development the *AtCDT1* gene is expressed. A Green-Fluorescence Protein (*GFP*) analysis will also be performed using the previously designed *pSB108* reporter construct (Michalopoulos et al. 2018; Appendix 1), which contains the *GFP* reporter gene driven by the *AtCDT1* promoter in the base plasmid *pMDC110* (Curtis & Grossniklaus 2003; Appendix 1). The *GFP* analysis will reveal where in the cell *AtCDT1* is localized. This information will help to confirm whether *AtCDT1* is localized in the cell membrane of root cells, and aids in binding up metal ions, as was predicted in Figure 1 and by previous research.

Since it has already been shown that *cdt1-2* mutant Arabidopsis plants overexpress *AtCDT1* in response to cadmium, but not aluminum (Figure 3), the question of what occurs when there is a lack of *AtCDT1* in response to aluminum or cadmium remains unanswered. To answer this question, a CRISPR-mediated knockout line of the *AtCDT1* gene of Arabidopsis plants was proposed and initiated in this research. The completion of the CRISPR construct, and its integration into Arabidopsis plants to create an *AtCDT1* knockout line, will be performed in the future, using the two preliminary constructs seen in Figure 11 of this paper.

The two preliminary constructs, containing either the exon 1 protospacer, or the upstream protospacer, would be used in a Gateway Cloning reaction to orient the two

protospacer sequences next to each other and then integrate them both into the final Cas9 plasmid. Gateway Cloning uses homologous recombination, which involves the crossing over of two sequences with high similarity, though they are of different origins (Figure 12). In this way, homologous recombination and Gateway Cloning are able to connect fragments of different plasmids, as long as these fragments are flanked by homologous sequences that can recombine with each other. The specific Gateway Cloning reaction that would be used in this case is the LR reaction, which involves recombination between the homologous *attL* and *attR* sites that flank the protospacer sequences. This LR reaction would occur twice: firstly, to align the protospacer sequences together in the same plasmid (Figure 13); and then to integrate both sequences into the Cas9-containing plasmid *pMR430* with the reporter gene *OLF-Citrine*, to create the final CRISPR construct (Figure 14; Appendix 1).

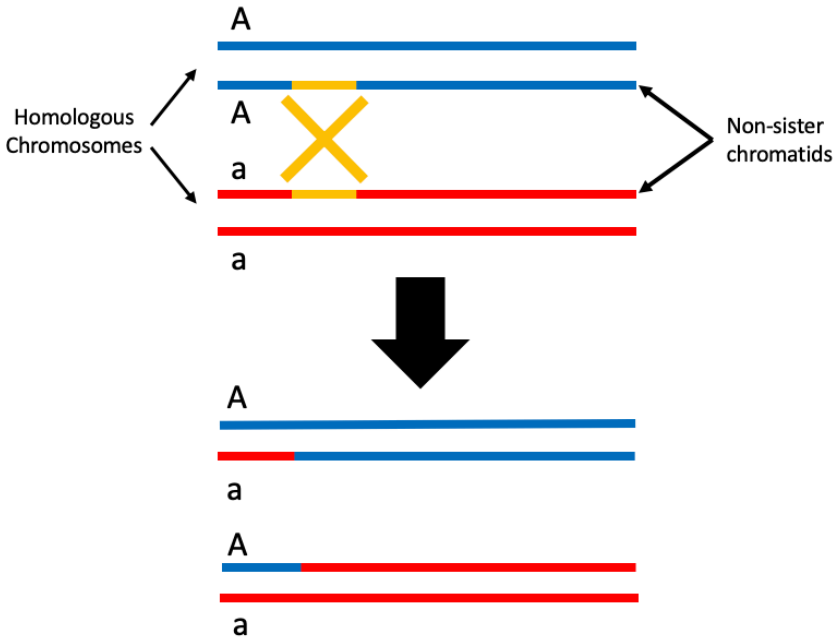

Figure 12: Example of homologous recombination between non-sister chromatids of homologous chromosomes in a crossing-over event during meiosis, resulting in the incorporation of a new sequence from a different source after recombination between two homologous sequences.

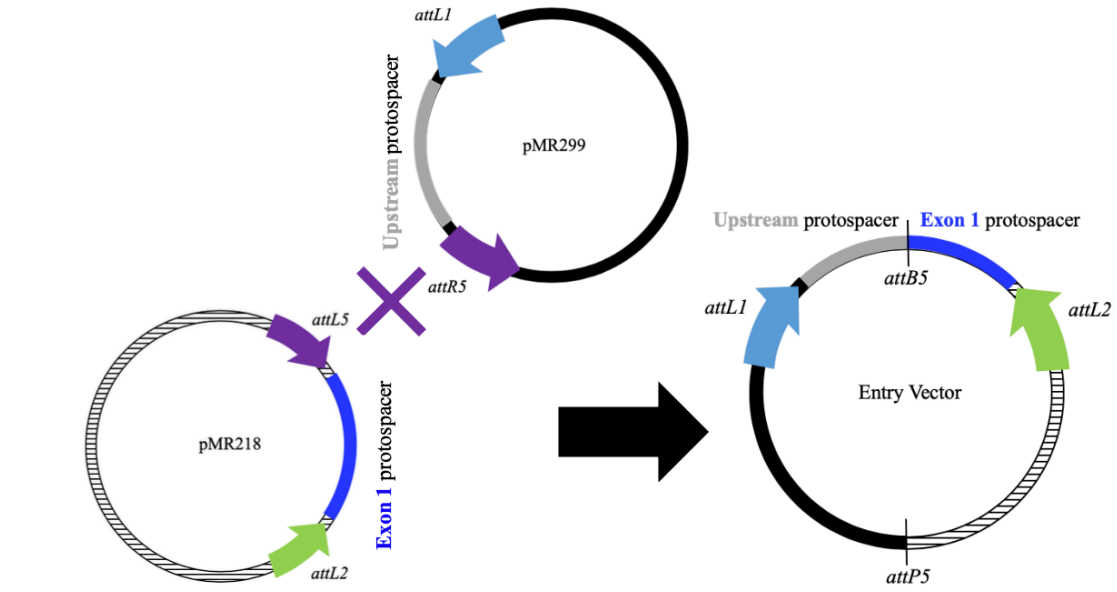

Figure 13: Diagram of first LR reaction, using homologous recombination between the *attL5* site of pMR218, containing the exon 1 protospacer, and the *attR5* site of pMR299, containing the upstream protospacer, to create an entry vector with the two protospacers adjacent to each other.

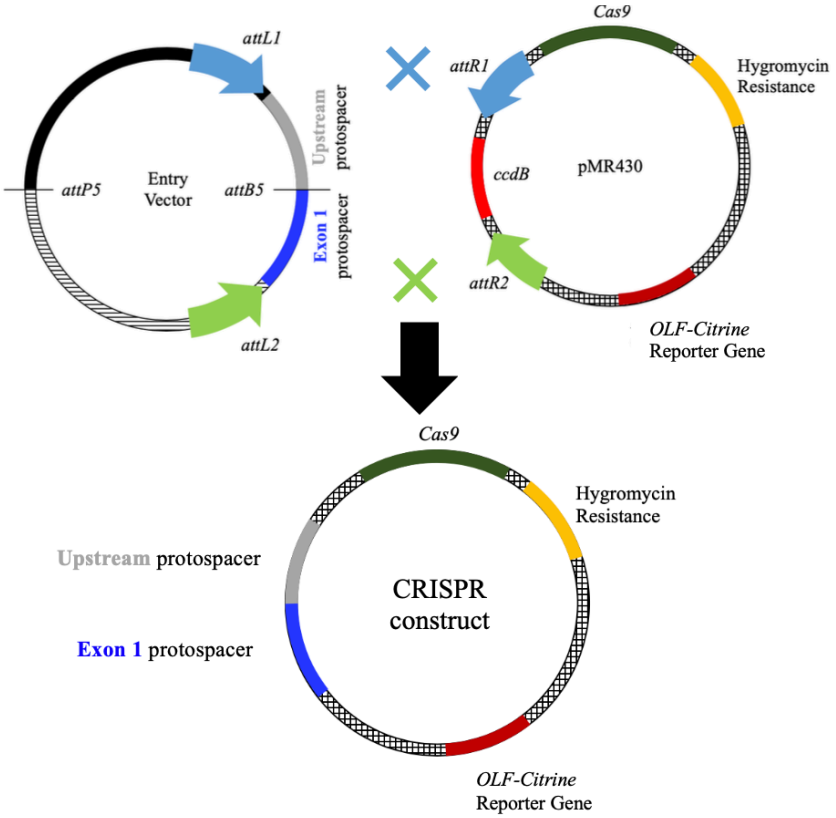

Figure 14: Diagram of the second LR reaction, using homologous recombination between the *attL1* and *attL2* sites of the entry vector (result of Fig. 13) and the *attR1* and *attR2* sites of pMR430, to create a CRISPR construct with the exon 1 protospacer and the upstream protospacer adjacent to each other as well as the *Cas9* sequence.

This final CRISPR construct would be amplified in *E. coli* and then sent for sequencing to confirm the presence of both of the protospacer sequences. The CRISPR construct would then be integrated into *A. tumefaciens* using Agrobacterium-mediated transformation (see Methods, above). The resulting *A. tumefaciens* cells containing the CRISPR construct would infect Arabidopsis plants to integrate the *Cas9*, protospacer, and reporter sequences into the Arabidopsis genome. These transgenic plants would transcribe the protospacer sequences, to become the guide RNA sequences, which would associate with the transcribed and translated Cas9 protein. This complex would then locate the *AtCDT1* gene and make double stranded cuts in the upstream region and exon 1 of the gene. The DNA would naturally try to repair itself using non-homologous end joining (NHEJ), which would induce such mutations as frameshift and nonsense mutations, thereby functionally destroying the *AtCDT1* gene.

The *AtCDT1* knockout line of Arabidopsis will be derived from these plants transformed with the CRISPR construct. This knockout line will be used in a number of experiments, involving increasing concentrations of aluminum or cadmium, in order to determine *AtCDT1*'s involvement in tolerance to either of these metals. The phenotypic representation of a lack of tolerance to these metals is an inhibition of root growth. To quantify this response, the root growth rate of each individual plant under either aluminum, cadmium, or normal treatment would be measured over the span of a set amount of time. If the knockout Arabidopsis plants lacking the *AtCDT1* gene show a decreased root growth rate in response to increasing concentrations of aluminum, in comparison to the wild type root growth rate under the same conditions, then this data would suggest that *AtCDT1* is important in the Arabidopsis response to aluminum toxicity. Similarly, if the *AtCDT1*

knockout plants show decreased root growth rate in response to increasing levels of cadmium, this data suggests that the gene is important in cadmium tolerance. There is the possibility that *AtCDT1* will be determined to be involved in both cadmium and aluminum tolerance, or neither.

The study of metal toxicity such as aluminum and cadmium toxicity, is important due to the commonality of naturally occurring aluminum in low pH soils, and the increasing commonality of pollution-enriched cadmium in low pH soils across the country and around the world. Studying genes involved in tolerance to these metals in different plants can reveal important mechanism used by plants to confer tolerance to metal toxicity, which can be used in molecular biology to increase crop yield in high metal, low pH soils.

Specifically, the CRISPR/Cas9 mechanism can be used, not to design knockouts as shown in this research, but to transgenically introduce genes important in tolerance to such metals as aluminum or cadmium into different plants, to promote better growth and increased crop yield in high metal, low pH soils. The use of CRISPR offers an interesting benefit, in that the introduced CRISPR/Cas9 sequences can be crossed out of the transgenic plant's genome, while maintaining the added metal tolerance gene, making it no longer a transgenic plant (Li 2018). The CRISPR/Cas9 technology would be inserted randomly into the plant's genome, and the introduced gene would also insert randomly into the genome in the following generation of seeds. Both the CRISPR/Cas9 technology and the introduced gene are heritable, but not necessarily on the same chromosomes, meaning that there would exist plants that carry the introduced gene, but not the CRISPR/Cas9 sequences (Li 2018).

The use of CRISPR also offers the benefit of rapid integration of such genes into plants, in as few as one generation after the initial transgenic plant (Li 2018). This rapid

integration differs from traditional breeding technologies that would require multiple crosses of plants with desired traits over multiple generations, before reaching the final plant carrying the desired genetic trait. The CRISPR/Cas9 mechanism is an important discovery in molecular biology and can be used to rapidly integrate transgenic sequences into organisms. CRISPR technology can specifically be used to integrate genes important in metal tolerance into plants, therefore allowing these plants to be grown in a wider range of soils, including those high in metal content and low in pH, which would help to increase crop yield of many plants that cannot naturally grow in such soils.

APPENDIX

1. Plasmid Glossary

Plasmid Name	Content	Original Citation
pMDC110	• *GFP* visual marker • Bacterial kanamycin resistance and plant hygromycin resistance selectable markers • *AscI* and *PacI* cut sites • *ccdb* negative selectable marker sequence	Curtis & Grossniklaus 2003
pMDC164	• *GUS* visual marker • Bacterial kanamycin resistance and plant hygromycin resistance selectable markers • *SalI* and *BamHI* cut sites • *ccdb* negative selectable marker sequence	Curtis & Grossniklaus 2003
pSP108	• *GFP* visual marker sequence driven by *AtCDT1* promoter • Bacterial kanamycin resistance and plant hygromycin resistance selectable markers	Michalopoulos et al. 2018
pSB109	• *GUS* visual marker sequence driven by *AtCDT1* promoter • Bacterial kanamycin resistance and plant hygromycin resistance selectable markers	Current study
pMR218	• *attL5* and *attL2* recombination sites • Bacterial kanamycin resistance selectable marker • *BbsI* cut sites	Ron 2018
pMR299	• *attL1* and *attR5* recombination sites • Bacterial kanamycin resistance selectable marker • *BbsI* cut sites	Ron 2018

Plasmid Name	Content	Original Citation
pMR430	• *Cas9* coding sequence • *attR1* and *attR2* sites • Bacterial kanamycin resistance and plant hygromycin resistance selectable markers • *OLF-Citrine* seed visual marker sequence • *ccdb* negative selectable marker sequence	Ron 2018
CRISPR construct	• Protospacer sequence targeting upstream region of *AtCDT1* gene • Protospacer sequence targeting exon 1 of *AtCDT1* gene • Bacterial kanamycin resistance and plant hygromycin resistance selectable markers • *OLF-Citrine seed visual marker sequence*	Not yet completed

2. Inoculation Media (IM): 50 mL
    a. 2.5 g sucrose
    b. 0.1 g MS
    c. 0.02 g MES
    d. pH 5.7
    e. Water to 50 mL
    f. 25 mL Silwet L-77 (add Silwet immediately before adding the IM to the *Agrobacterium*)

Literature Cited:

1. The Arabidopsis Information Resource. 2020. "AT1G52827(CDT1)." <https://www.arabidopsis.org/servlets/TairObject?id=1000530066&type=locus>

2. The Broad Institute. 2019. "Research Highlights: CRISPR." <https://www.broadinstitute.org/research-highlights-crispr>

3. The Broad Institute. 2018. "Questions and Answers about CRISPR." <https://www.broadinstitute.org/what-broad/areas-focus/project-spotlight/questions-and-answers-about-crispr>

4. Clemens S. 2001. Molecular mechanisms of plant metal tolerance and homeostasis. *Planta*. 212: 475-486.

5. Clough SJ and Bent AF. 1998. Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *The Plant J.* **6(6)**: 735–743.

6. Curtis MD and Grossniklaus U. 2003. A Gateway Cloning Vector Set for High-Throughput Functional Analysis of Genes in Planta. *Plant Physiol.* **133** (2) 462-469.

7. Eekhout T, Larsen P, and De Veylder L. 2017. Modification of DNA Checkpoints to Confer Aluminum Tolerance. *Trends Plant Sci.* **2**: 102-105.

8. Gazey C. 2018. "Effects of Soil Acidity." *Agriculture and Food. Department of Primary Industries and Regional Development, Government of Western Australia.* <https://agric.wa.gov.au/n/2113>

9. Heck and Bailey. 1950. Chelation of trace metals in nutrient solutions. *Plant Physiol.* **4:** 573-582.

10. Hoekenga OA, Maron LG, Piñeros MA, Cançado GM, Shaff J, Kobayashi Y, Ryan PR, Dong B, Delhaize E, Sasaki T, Matsumoto H, Yamamoto Y, Koyama H, and Kochian LV. 2006. AtALMT1,which encodes a malate transporter, is identified as one of several genes critical for aluminum tolerance in *Arabidopsis. Proc. Natl. Acad. Sci. U.S.A.* **103**: 9738–9743.

11. Iuchi S, Koyama H, Iuchi A, Kobayashi Y, Kitabayashi S, Kobayashi Y, Ikka T, Hirayama T, Shinozaki K, and Kobayashi M. 2007. Zinc finger protein STOP1 is critical for proton tolerance in *Arabidopsis* and co-regulates a key gene in aluminum tolerance. *Proc. Natl. Acad. Sci. U.S.A.* **104**: 9900–9905.

12. Kuramata M, Masuya S, Takahashi Y, Kitagawa E, Inoue C, Ishikawa S, Youssefian S, and Kusano T. 2009. Novel cysteine-rich peptides from Digitaria ciliaris and Oryza sativa enhance tolerance to cadmium by limiting its cellular accumulation. *Plant Cell Physiol*. **50**: 106-117.

13. Larsen PB, Geisler MJ, Jones CA, Williams KM, and Cancel JD. 2005. *ALS3* encodes a phloem-localized ABC transporter-like protein that is required for aluminum tolerance in Arabidopsis. *Plant J.* **41**: 353–363.

14. Li, Y. 2018. "These CRISPR-modified crops don't count as GMOs." *The Conversation*. < https://theconversation.com/these-crispr-modified-crops-dont-count-as-gmos-96002>

15. Michopoulos C, Eldredge C, and Bush S. 2019. Determining the Role of CDT1 Under Metal Stress in Arabidopsis. Trinity College, Hartford, CT.

16. Pauwels L, De Clercq R, Goossens J, Iñigo S, Williams C, Ron M, Britt A, and Goossens A. 2018. A Dual sgRNA Approach for Functional Genomics in *Arabidopsis thaliana*. *G3*. **8**: 2603–2615.

17. Ron, M. 2018. Cloning of Custom CRISPRs: All the new vectors are modified from the pDe_CAS9. *University of California, Davis*.

18. Vector BioLabs. "CRISPR/Cas9 AAV and Adenovirus Product Catalog." <https://www.vectorbiolabs.com/cas9-aav-adenovirus/>

19. Xia J, Yamaji N, and Ma JF. 2013. A plasma membrane-localized small peptide is involved in rice aluminum tolerance. *Plant J.* **76**: 345-355.

20. Xu Y, Yu Z, Zhang D, Huang J, Wu C, Yang G, Yan K, Zhang S, and Zheng C. 2017. CYSTM, a Novel Non-Secreted Cysteine-Rich Peptide Family, Involved in Environmental Stresses in *Arabidopsis thaliana*. *Plant Cell Physiol.* **2:** 423-428.